Mouse matriptase-2: identification, characterization and comparative mRNA expression analysis with mouse hepsin in adult and embryonic tissues

John D. HOOPER[1], Luisa CAMPAGNOLO, Goodarz GOODARZI, Tony N. TRUONG, Heidi STUHLMANN and James P. QUIGLEY[2]

Division of Vascular Biology, Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, U.S.A.

We report the identification and characterization of mouse matriptase-2 (m-matriptase-2), an 811-amino-acid protein composed of an N-terminal cytoplasmic domain, a membrane-spanning domain, two CUB (complement protein subcomponents C1r/C1s, urchin embryonic growth factor and bone morphogenetic protein 1) domains, three LDLR (low-density-lipoprotein receptor class A) domains and a C-terminal serine-protease domain. All m-matriptase-2 protein domain boundaries corresponded with intron/exon junctions of the encoding gene, which spans approx. 29 kb and comprises 18 exons. Matriptase-2 is highly conserved in human, mouse and rat, with the rat matriptase-2 gene (*r-maltriptase-2*) predicted to encode transmembrane and soluble isoforms. Western-blot analysis indicated that m-matriptase-2 migrates close to its theoretical molecular mass of 91 kDa, and immunofluorescence analysis was consistent with the proposed surface membrane localization of this protein. Reverse-transcription PCR and *in-situ*-hybridization analysis indicated that m-matriptase-2 expression overlaps with the distribution of mouse hepsin (m-hepsin, a cell-surface serine protease identified in hepatoma cells) in adult tissues and during embryonic development. In adult tissues both are expressed at highest levels in liver, kidney and uterus. During embryogenesis m-matriptase-2 expression peaked between days 12.5 and 15.5. m-hepsin expression was biphasic, with peaks at day 7.5 to 8.5 and again between days 12.5 and 15.5. *In situ* hybridization of embryonic tissues indicated abundant expression of both m-matriptase-2 and m-hepsin in the developing liver and at lower levels in developing pharyngo–tympanic tubes. While m-hepsin was detected in the residual embryonic yolk sac and with lower intensity in lung, heart, gastrointestinal tract, developing kidney tubules and epithelium of the oral cavity, m-matriptase-2 was absent in these tissues, but strongly expressed within the nasal cavity by olfactory epithelial cells. Mechanistic insight into the potential role of this new transmembrane serine protease is provided by its novel expression profile in embryonic and adult mouse.

Key words: hepsin, membrane, mouse matriptase-2, serine protease.

INTRODUCTION

The type II transmembrane serine proteases (TTSPs) are a family of proteolytic enzymes characterized by a short N-terminal cytoplasmic tail, a membrane-spanning region, potential ligand-binding domains and a C-terminal trypsin-like serine-protease domain [1]. Member proteases include enteropeptidase, the activator of trypsinogen in the digestive enzyme cascade [2], corin, a heart-specific [3,4] pro-(atrial natriuretic peptide) (pro-ANP) activator [5,6], hepsin (a cell-surface serine protease identified in hepatoma cells) which has consistently been shown to be highly up-regulated at the mRNA level in prostate cancer [7–12], and matriptase/membrane-type serine protease 1 (MT-SP1), which is widely expressed in normal epithelial tissues [13,14] as well as tumours of epithelial origin [15–17]. Matriptase/MT-SP1 has the ability to activate urokinase-type plasminogen activator, protease-activated receptor 2 and hepatocyte growth factor [18,19], and the crystal structure of the catalytic domain of this TTSP in the presence of benzamidine and in complex with bovine pancreatic trypsin inhibitor has recently been reported [20]. The functional significance of this enzyme in epidermal barrier function, hair-follicle development and thymic homoeostasis has recently been demonstrated by the generation of matriptase/MT-SP1$^{-/-}$ mice [21]. Other human members of the TTSP family include human airway trypsin-like serine protease (HAT) [22], MSPL [23], serine protease DESC1 [24], TMPRSS2 ('transmembrane protease, serine 2') [25], TMPRSS3 [26], TMPRSS4 (formerly designated TMPRSS3) [27] and spinesin/TMPRSS5 [28]. Mouse orthologues have been described for enteropeptidase [29], matriptase/MT-SP1 [30], corin [31], hepsin [32] and TMPRSS2 [33].

In addition to a role at the cell surface, several of the TTSPs have been described as soluble proteins. Matriptase/MT-SP1 has been isolated from breast-cancer-cell-conditioned medium [34], as well as from human breast milk [35]. In mouse, post-translational processing is required for release of matriptase/MT-SP1 from the plasma membrane [36]. In contrast with mouse matriptase (m-matriptase)/MT-SP1, a soluble form of MSPL is produced as a result of transcriptional mechanisms which produce an mRNA encoding a protein without a transmembrane domain [23]. Enteropeptidase is produced in the duodenum by enterocytes, and a soluble form of the protein is present in mucinous secretions of

Abbreviations used: AEBSF, 4-(2-aminoethyl)benzenesulphonyl fluoride hydrochloride; ANP, atrial natriuretic peptide; CHO, Chinese-hamster ovary; CUB, complement protein subcomponents C1r/C1s, urchin embryonic growth factor and bone morphogenetic protein 1; EST, expressed sequence tag; FBS, fetal-bovine serum; HAT, human airway trypsin-like serine protease; LDLR, low-density-lipoprotein receptor class A; RT, reverse transcription; TTSP, type II transmembrane serine protease; MT-SP1, membrane-type serine protease 1; m-hepsin, mouse hepsin; h-, m- and r-matriptase, human, mouse and rat matriptase respectively; A.T.C.C., American Type Culture Collection; FLAG epitope, DYKDDDDK (in one-letter code); MYC epitope, EQKLISEEDL; mAb, monoclonal antibody; dpc, days *post coitus*; PSP-I, porcine seminal plasma-1.

[1] Present address: Centre for Molecular Biotechnology, Queensland University of Technology, Brisbane, Queensland, Australia.

[2] To whom correspondence should be addressed (e-mail jquigley@scripps.edu).

The nucleotide sequences reported in this paper have been deposited in the DDBJ/GenBank®/EMBL and GSDB Nucleotide Sequence Databases under the accession numbers AY240929, BK000520, AY055383, AY055384 and AY234104.

the small intestine [37]. A soluble form of TMPRSS2 is detected in the media of cultured prostate-cancer cells, as well as in the sera of prostate tumour-bearing mice [38]. Following synthesis and transport to the cell surface, HAT is released from tracheal serous glands as part of the host immune defence system [22].

The most recently reported human TTSP, matriptase-2, was named on the basis of its structural similarity to matriptase/MT-SP1 [39]. The recombinant matriptase-2 catalytic domain cleaves following arginine residues and is inhibited by the serine protease inhibitors PMSF, 4-(2-aminoethyl)benzenesulphonyl fluoride hydrochloride (AEBSF), leupeptin and aprotinin. This domain also demonstrated *in vitro* hydrolytic activity against a number of matrix and basement-membrane components [39]. Here we report the identification of the mouse orthologue of matriptase-2 (m-matriptase), the genomic structure of the encoding gene, partial protein characterization and mRNA distribution in mouse adult and embryonic tissues. We also describe potential membrane-bound and soluble isoforms of rat matriptase-2 (r-matriptase-2). A novel mouse hepsin (m-hepsin) splice variant from brain was also identified during comparative analysis of m-matriptase-2 and m-hepsin mRNA expression.

EXPERIMENTAL

Database mining/bioinformatics

The Ensembl database (www.ensembl.org) was searched for S1-serine-protease-domain-encoding genes using the Interpro domain number for trypsin (S1)-like serine proteases IPR001254. Nucleotide and amino acid sequences were used to search GenBank® databases using algorithms available at the National Center for Biotechnology Information (NCBI) website. The m-matriptase-2 amino acid sequence was analysed for structural domains, cellular processing signals and consensus post-translational modification motifs using the Prosite database [40], the SMART algorithm [41], the PSORT algorithm [42] and the NetPhos 2.0 algorithm [43]. Protein sequences were aligned using EclustalW at the Australian National Genome Information website.

Cell culture

Human cervical adenocarcinoma, HeLa and Chinese-hamster ovary (CHO)-K1 cells were obtained from the American Type Culture Collection (A.T.C.C., Manassas, VA, U.S.A.), maintained as monolayer cultures in Dulbecco's modified Eagle's medium (Invitrogen, Carlsbad, CA, U.S.A.) or Ham's F12K medium (A.T.C.C.) respectively, supplemented with 10 % (v/v) fetal-bovine serum (FBS) (HyClone, Logan, UT, U.S.A.), sodium pyruvate, penicillin/streptomycin and non-essential amino acids (Invitrogen) and grown in a humidified 5 % CO_2 atmosphere at 37 °C.

DNA constructs

m-matriptase-2 (GenBank® accession number BF385919) and m-hepsin (GenBank® accession number AI528314) expressed sequence tag (EST) clones were purchased (Invitrogen). A construct designated m-matriptase-2-flag, in pCMV-SPORT6 expressing m-matriptase-2 tagged at the C-terminus with a FLAG (DYKDDDDK) epitope, was generated by PCR. A construct in pcDNA3 designated m-hepsin-myc, expressing m-hepsin tagged at the C-terminus with a MYC (EQKLISEEDL) epitope, was also generated by PCR. In each case the 3′ PCR primer incorporated sequence encoding the respective tag immediately before the stop codon. Constructs used to generate probes for *in situ* hybridization were produced by cloning PCR products into the pCRII-TOPO vector (Invitrogen). For m-matriptase-2, PCR was performed with primers F7 (CCCTCTCTGGACTACGGCTTGGC) and R10 (CGGCTCACCTTGAAGGACAC). For m-hepsin PCR was performed with primers F1 (CGCACTCGGAGCTGGATGTG) and R2 (CTGTCGCCCTGGCACGCATC). All constructs were sequenced to confirm that no errors had been introduced during cloning.

Transient transfections and Western-blot analysis

Cells were transiently transfected with the m-matriptase-2-flag expression construct using PolyFect® reagent (Qiagen, Valencia, CA, U.S.A.) as described by the manufacturer. After incubation for 48 h at 37 °C, cells were lysed in ice-cold buffer containing 10 mM Tris (pH 8.0), 150 mM NaCl, 1 % Triton X-100, 5 mM EDTA and 1 × Complete Mini, EDTA-Free Protease Inhibitor Cocktail (Roche, Indianapolis, IN, U.S.A.). Insoluble material was removed by centrifugation at 15 800 ***g*** for 10 min. Lysates were separated by electrophoresis through SDS/polyacrylamide gels under reducing conditions, then transferred to nitrocellulose membranes (Millipore). Membranes were blocked in 5 % (w/v) non-fat skim milk in PBS containing 0.1 % Tween 20, then incubated overnight at 4 °C with anti-FLAG M2 monoclonal antibody (mAb; 0.8 μg/ml; Sigma, St. Louis, MO, U.S.A.). After they had been washed, membranes were incubated for 2 h at room temperature with goat anti-mouse IgG (0.16 μg/ml; Pierce, Rockford, IL, U.S.A.) and immunoreactive bands detected by enhanced chemiluminescence (Pierce, Rockford, IL, U.S.A.).

Immunofluoresence

Cells plated on coverslips were transiently transfected with either the m-matriptase-2-flag or m-hepsin-myc expression constructs. After incubation for 48 h at 37 °C, cells were washed with PBS, fixed in 2 % formaldehyde, then blocked with 5 % normal goat serum in PBS. Following a 1 h incubation at 37 °C with either anti-FLAG M2 mAb (2 μg/ml) or anti-MYC 9B11 mAb (2 μg/ml; Cell Signalling Technology, Beverly, MA, U.S.A.) in blocking buffer, cells were washed with PBS, then incubated with FITC-conjugated goat anti-mouse IgG (2 μg/ml; Jackson ImmunoResearch Laboratories, Inc., West Grove, PA, U.S.A.). Labelled cells were visualized and photographed using a Bio-Rad 1024 MRC2 scanning confocal imaging system.

Reverse transcription (RT)-PCR

CD1 staged embryos (morning of vaginal plug corrected as 0.5 day *post coitus*; dpc) and adult tissues were collected, snap-frozen in liquid nitrogen, then stored at − 80 °C. Tissues and embryos were homogenized in the presence of either Trizol reagent (Invitrogen) or a solution containing 4 M guanidinium thiocyanate, 25 mM sodium citrate, pH 7, 0.5 % sodium sarcosyl and 0.1 M 2-mercaptoethanol. Following removal of insoluble material and extraction, total RNA was precipitated in the presence of sodium acetate and ethanol. RT was performed on 1 μg of total RNA using Superscript II (Invitrogen). PCR was performed with Platinum *Taq* DNA polymerase (Invitrogen) on 1 μl of the resulting cDNA using primer pairs F7/R10 (m-matriptase-2) or F1/R2 (m-hepsin). cDNA quality was assessed using β-actin primers (26 cycles of PCR). PCR cycling conditions: m-matriptase-2, 94 °C for 3 min, 35 cycles of 94 °C for 30 s, 58 °C for 30 s and 72 °C for 50 s,

followed by a final 72 °C extension for 10 min; m-hepsin, 94 °C for 3 min, 32 cycles of 94 °C for 30 s, 60 °C for 30 s and 72 °C for 45 s, followed by a final 72 °C extension for 10 min. Reaction products were analysed by agarose-gel electrophoresis and photographed. PCR products from adult liver (m-matriptase-2), adult brain (m-hepsin), adult spleen (m-hepsin) and 7.5-day embryo (m-hepsin) were cloned into the vector pCRII-TOPO and sequenced. Density profiling of PCR product bands was performed using Scion Image software (Scion Corporation, Frederick, MA, U.S.A.).

In situ hybridization

m-matriptase-2 sense and antisense ^{35}S-labelled cRNA riboprobes were generated using SP6 and T7 RNA polymerase (Roche) following *Xho*I and *Spe*I digestion respectively of the F7/R10 construct described above. m-hepsin sense and antisense ^{35}S-labelled cRNA riboprobes were generated using T7 and SP6 RNA polymerase following *Spe*I and *Xho*I digestion respectively of the F1/R2 construct described above. Following collection, CD1 staged mouse embryos and adult tissues were immediately fixed by overnight incubation at 4 °C in freshly prepared 4 % paraformaldehyde, then dehydrated and embedded in paraffin at 60 °C. *In situ* hybridization using ^{35}S-radiolabelled probes was performed on deparaffinized and Proteinase K (20 μg/ml)-treated sections (5 μm thick) as previously described [44,45]. The final concentration for each antisense and sense riboprobe was 35 c.p.m./ml in hybridization buffer. Following riboprobe incubation at 52 °C for 16 h and several washes of increasing stringency, slides were dipped in Kodak NBT-2 photoemulsion, dried for 1 h, then exposed for 5 days at 4 °C. Adult tissue sections were counterstained with haematoxylin and eosin. Photographs were taken using a Zeiss Axioplan 2 microscope.

RESULTS

Identification and characterization of the m-matriptase-2 cDNA and gene (*m-matriptase-2*)

A search of the Ensembl database was performed to identify novel TTSPs. Five unannotated human candidate proteins were identified by searching with the Interpro domain number for trypsin (S1)-like serine proteases (IPR001254). The nucleotide sequence encoding one of these partial proteins, the product of the predicted gene entry ENSG00000133420, when searched against the GenBank® EST database closely matched an unpublished murine EST entry (BF385919). The corresponding clone, which was generated from mouse liver, was obtained and completely sequenced. The complete cDNA and protein sequence is shown in Figure 1. Analysis of the deduced amino acid sequence indicated that the protein has the features of a TTSP. These include the lack of a consensus secretory signal peptide, a potential transmembrane domain delineating a cytoplasmic tail of 59 residues, a region containing potential ligand binding domains and a C-terminal trypsin-like serine-protease domain (Figure 2A). Consistent with the naming of recently identified members of the TTSP family [1,28], we initially designated the encoded protein mouse TMPRSS6. However, as the human orthologue was subsequently reported as matriptase-2 [39], here we refer to the protein as m-matriptase-2.

The longest open reading frame of the m-matriptase-2 cDNA spans nucleotides 187–2619 and encodes a protein of 811 amino acids with a theoretical molecular mass of 91.0 kDa. The cDNA contains 547 bp of 3′-untranslated sequence, including a consensus polyadenylation signal (AATAAA) at position 3139. The cytoplasmic tail of the deduced protein contains consensus protein kinase C phosphorylation sites at Ser^{10} and Thr^{47}. Analysis of the region between the transmembrane and catalytic domains identified two CUB (complement protein subcomponents C1r/C1s, urchin embryonic growth factor and bone morphogenetic protein 1) domains (residues 221–330 and 335–452) and three LDLR (low density lipoprotein receptor-like) domains (residues 457–490, 490–526 and 530–567) (Figures 1 and 2A). Seven consensus N-glycosylation sites were detected in the extracellular portion of the protein; none of these was within the catalytic domain (Figure 1). The proteolytic domain encompasses amino acids 577–811, with the catalytic triad located at His^{617}, Asp^{668} and Ser^{762}. This region and the activation domain contain ten highly conserved cysteine residues predicted to form the following disulphide-bonded residues: 568–688, 602–618, 702–768, 733–747 and 758–787. As the first disulphide bond links the pro and catalytic domains, following activation at the conserved Arg^{576}–Ile^{577} bond the serine-protease domain is predicted to remain linked to the activation domain. On the basis of the presence of Asp^{756} at the bottom of the substrate-binding pocket, which in the linear protein sequence is located six residues before the catalytic serine residue, m-matriptase-2 is predicted to have trypsin-like specificity with cleavage following arginine or lysine residues. A conserved Ser^{782}-Trp-Gly motif is predicted to be located at the top of the substrate-binding pocket, permitting correct orientation of the scissile bond of the substrate.

Alignment of the m-matriptase-2 cDNA against the mouse genome localized the gene to PAC clone RP23-79F10 generated from mouse chromosome 15E2, which is syntenic with the location of the human matriptase (h-matriptase-2) gene (*h-matriptase-2*) at 22q13.1. The *m-matriptase-2* gene spans approx. 29 kb with 18 exons and 17 intervening introns (Figure 2A). Exon sizes ranged from 42 to 706 bp and introns from 206 to 5667 bp. All m-matriptase-2 protein domain boundaries corresponded with intron/exon junctions of the encoding gene (Figure 2A). The initiating methionine codon is located within exon 2, which encodes the complete cytoplasmic and transmembrane domains of m-matriptase-2. The first CUB domain is encoded by two exons (7 and 8), and the second CUB domain by three exons (9–11) The three LDLR domains of matriptase-2 are each encoded by separate exons (12, 13 and 14). The serine-protease domain, including the pro region, is encoded by four exons (15–18). Sequences at the 5′ and 3′ end of each intron conform to the GT-AG rule for splice-site recognition (Figure 2B) [46].

Matriptase-2 is conserved in human, mouse and rat, and the *r-matriptase-2* gene is predicted to encode transmembrane and soluble isoforms

Very recently, an unpublished GenBank entry (accession number XP_235768) has been released that encodes a rat protein with high sequence identity with matriptase-2. An alignment of this protein against h-matriptase-2 and m-matriptase-2 is shown in Figure 3(A). Human and mouse proteins share 83.4 % identity (676/811), mouse and rat 85.2 % (691/811) and human and rat 77.1 % identity (625/811). Each protein contains two CUB domains, three LDLR domains and a serine-protease domain. The seven consensus N-glycosylation sites within m-matriptase-2 are absolutely conserved in h- and r-matriptase-2. In addition to the ten cysteine residues within the serine-protease and activation domains, there are 27 other cysteine residues within the m-matriptase-2 extracellular domain which are conserved in human and rat. On the basis of the crystal structure of the spermadhesin

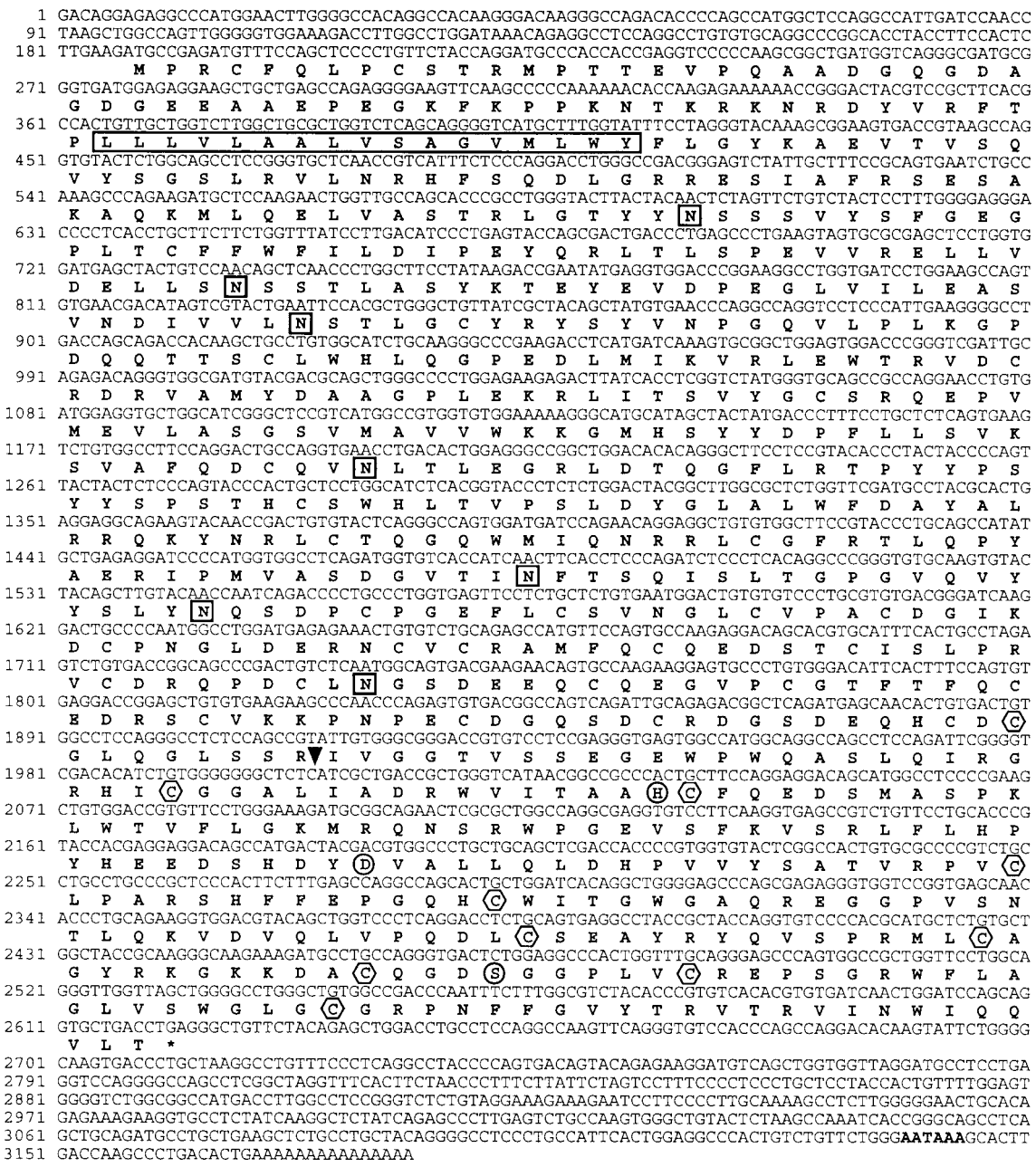
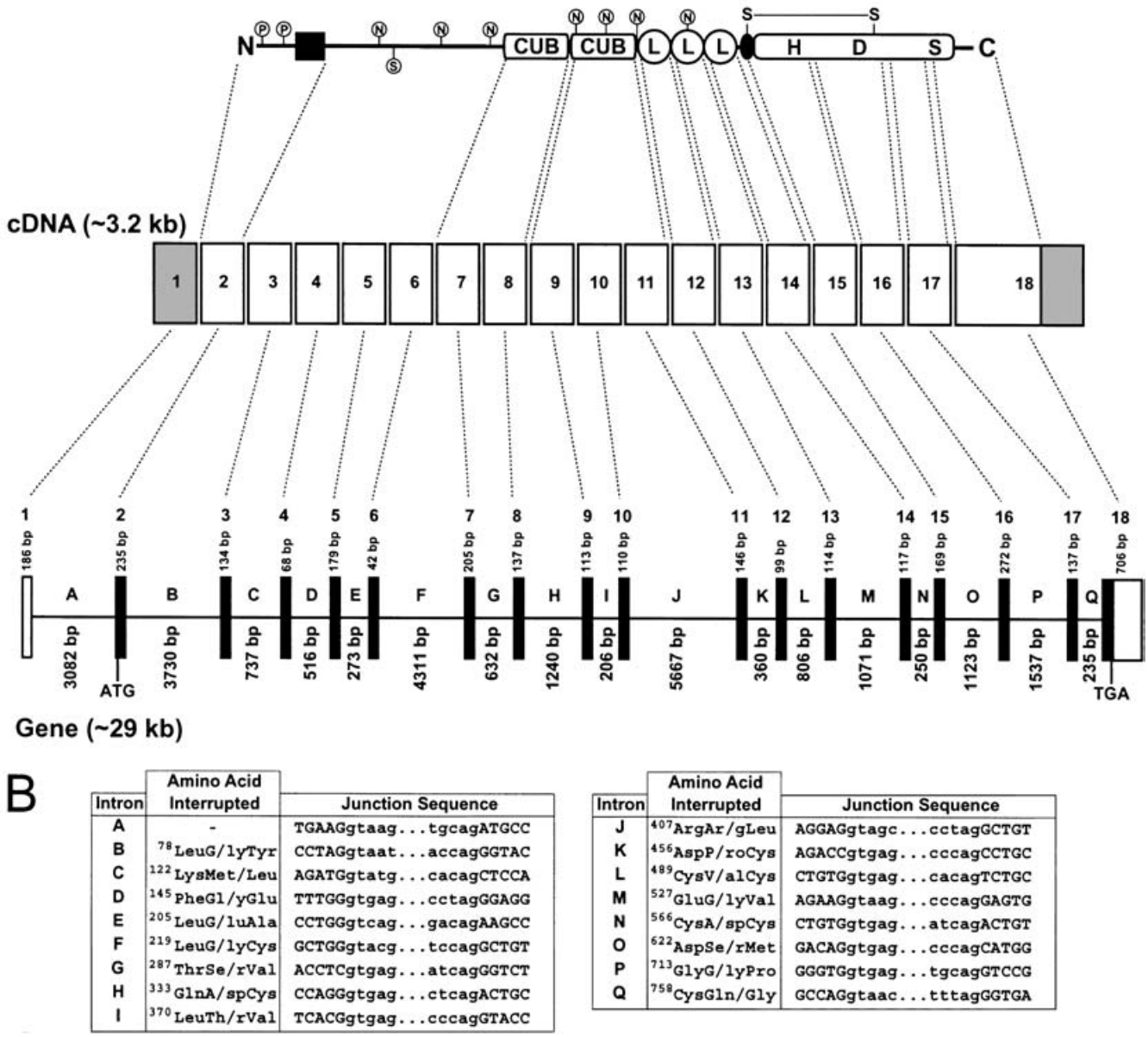


Figure 1 Nucleotide and deduced amino acid sequence of m-matriptase-2

Predicted catalytic histidine, aspartate and serine residues are circled. The zymogen activation site is indicated by an arrowhead. Cysteine residues within the activation and catalytic domains predicted to form disulphide bonds are surrounded by hexagons. Potential N-glycosylation sites are surrounded by squares, as is the transmembrane domain. The polyadenylation signal is in **bold**.

PSP-I (porcine seminal plasma-1)–PSP-II heterodimer [47], eight of these cysteine residues are predicted to form two disulphide links within each of the CUB domains. Similarly, on the basis of the structure of the LDLR extracellular domain [48], the six cysteine residues of each of the m-matriptase-2 LDLR domains are predicted to form three disulphide bonds. Accordingly, Cys^{152} of m-matriptase-2, and the corresponding residue within the human and rat proteins, is unpaired intramolecularly and therefore has the potential to form homo- or hetero- linkages.

The r-matriptase-2 amino acid sequence is divergent from the human and mouse proteins at the N-terminus and at the border of the first and second CUB domains, where the rat protein contains an additional 17 amino acids (Figure 3A). In contrast with h- and m-matriptase-2, the rat protein lacks a predicted membrane-spanning domain. Instead, r-matriptase-2 contains a hydrophobic region at amino acids 1–18 which conforms to the consensus for an N-terminal secretory signal peptide [49]. These variations at the protein level between m- and r-matriptase-2 begin to occur exactly at the junctions of exons 2/3 and exons 8/9 respectively of the mouse gene. To examine the possibility that these species differences have occurred as a result of mRNA splice variants, we analysed rat genomic DNA obtained from the GenBank® database to identify the *r-matriptase-2* exons encoding the divergent protein regions, and



B

Intron	Amino Acid Interrupted	Junction Sequence
A	-	TGAAGgtaag...tgcagATGCC
B	^{78}LeuG/lyTyr	CCTAGgtaat...accagGGTAC
C	^{122}LysMet/Leu	AGATGgtatg...cacagCTCCA
D	^{145}PheGl/yGlu	TTTGGgtgag...cctagGGAGG
E	^{205}LeuG/luAla	CCTGGgtcag...gacagAAGCC
F	^{219}LeuG/lyCys	GCTGGgtacg...tccagGCTGT
G	^{287}ThrSe/rVal	ACCTCgtgag...atcagGGTCT
H	^{333}GlnA/spCys	CCAGGgtgag...ctcagACTGC
I	^{370}LeuTh/rVal	TCACGgtgag...cccagGTACC

Intron	Amino Acid Interrupted	Junction Sequence
J	^{407}ArgAr/gLeu	AGGAGgtagc...cctagGCTGT
K	^{456}AspP/roCys	AGACCgtgag...cccagCCTGC
L	^{489}CysV/alCys	CTGTGgtgag...cacagTCTGC
M	^{527}GluG/lyVal	AGAAGgtaag...cccagGAGTG
N	^{566}CysA/spCys	CTGTGgtgag...atcagACTGT
O	^{622}AspSe/rMet	GACAGgtgag...cccagCATGG
P	^{713}GlyG/lyPro	GGGTGgtgag...tgcagGTCCG
Q	^{758}CysGln/Gly	GCCAGgtaac...tttagGGTGA

Figure 2 m-matriptase-2 protein and gene structure

(**A**) Upper panel: protein structural features marked include transmembrane (black box), two CUB, three LDLR (L), activation (black oval) and proteolytic (boxed HDS) domains. Consensus phosphorylation (P) and N-glycosylation (N) sites are marked. The unpaired Cys^{152} is indicated (S). The conserved disulphide bond linking the activation and catalytic domains is indicated. Middle panel: the m-matriptase-2 cDNA is represented by its constituent exons (numbered 1–18). Non-coding exon regions are shown in grey. Dotted lines indicate exons encoding the protein structural domains in the upper panel. Lower panel: Genomic organization. Exons are represented by numbered boxes with exon size indicated below. Exon coding regions are filled. Introns are represented by letters (A–Q) with intron size indicated. The Figure is not drawn to scale. (**B**) Sequence at m-matriptase-2 exon/intron junctions. The exon sequence is in upper-case and intron sequence in lower-case letters. The phase of the codon interrupted by each intron is indicated within the corresponding encoded amino acid by a forward stroke.

also to determine whether there is any potential for the rat genome to encode a transmembrane r-matriptase-2 isoform. This analysis identified rat chromosome 7 contig NW_044075 as containing the *r-matriptase-2* gene. As shown in Figure 3(B) (bottom of panel), two exons, of 52 and 108 bp, encode the variant N-terminus of r-matriptase-2. Also identified was a 51 bp exon encoding the 17 additional amino acids located between the first and second CUB domains of r-matriptase-2 (results not shown). In addition, by using the GenBank® Blast2 algorithm to align the m-matriptase-2 N-terminus against contig NW_044075, we were able to identify a 235 bp putative exon, located between the identified second and third coding exons of *r-matriptase-2*, which encodes an N-terminus highly homologous with the h- and m-matriptase-2 cytoplasmic and transmembrane domains (Figure 3B, top of panel). This potential exon is 89 % identical with, and has the same codon interrupted by the following intron

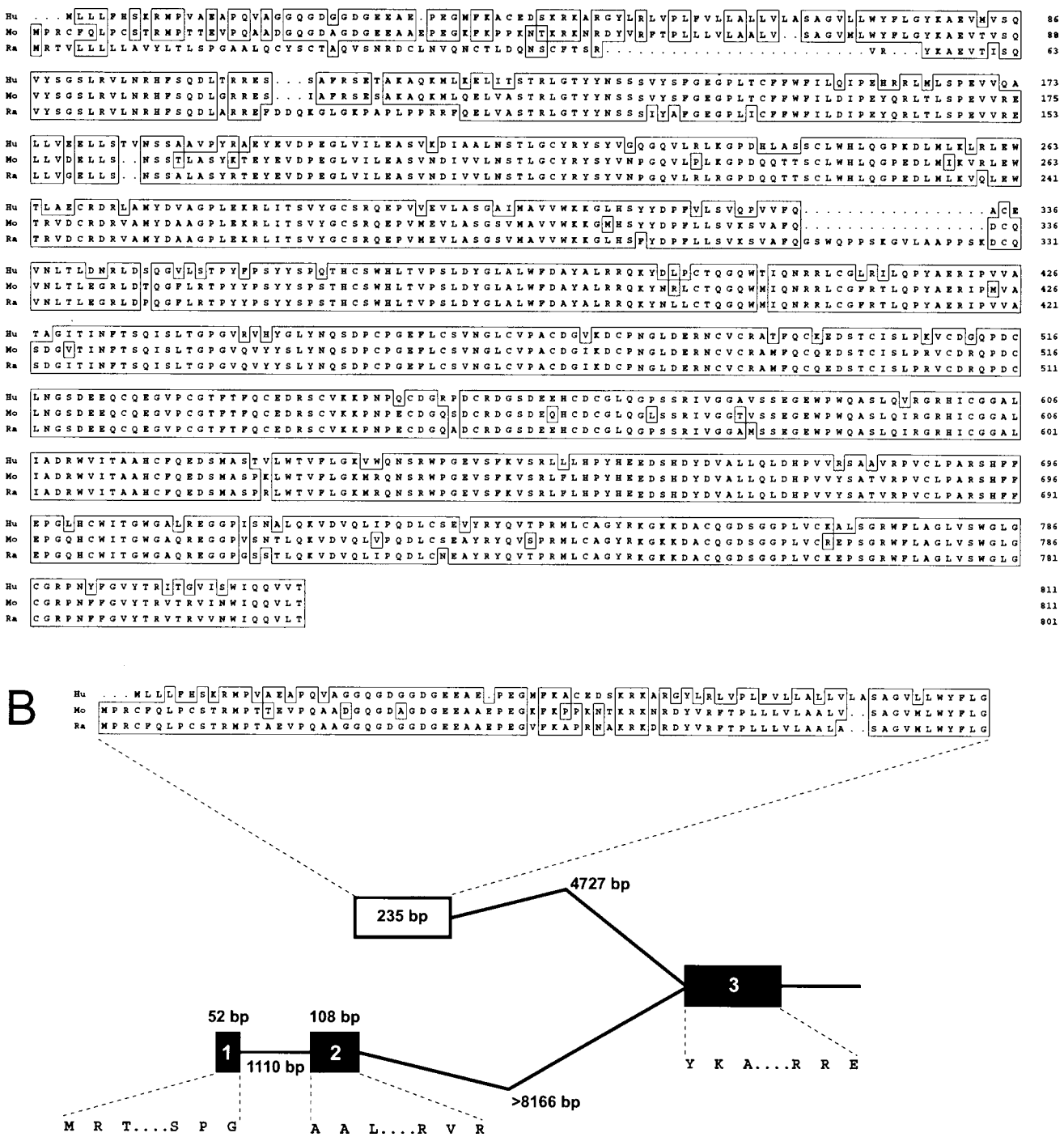


Figure 3 Matriptase-2 interspecies comparison

(**A**) Sequence alignment of human (Hu), mouse (Mo) and rat (Ra) matriptase-2 amino acid sequences. Residues which are identical between at least two of the sequences are boxed. GenBank® accession numbers: human, AY055384; mouse, AY240929; rat, XP_235768. Note that the sequence for h-matriptase-2 comes from our GenBank® entry, which is nine residues longer at the N-terminus than the sequence published by Velasco et al. [39]. (**B**) *r-matriptase-2* 5′ exons. Filled boxes represent exons identified by aligning the GenBank® r-matriptase-2 cDNA (accession number XM_235768) against rat genomic contig NW_044075. The unfilled box represents a putative exon identified by aligning the m-matriptase-2 N-terminus against contig NW_044075. Protein sequence encoded by this putative *r-matriptase* exon is aligned against the corresponding regions of h- and m-matriptase-2 at the top of the panel. Exon and intron sizes are marked. The residues at the boundaries of each exon are indicated.

and in the same phase as, *m-matriptase-2* exon 2. The 5′ splice site of the following *r-matriptase-2* intron commences with a GT dinucleotide and therefore conforms with the GT-AG rule for splice-site recognition. Thus the *r-matriptase-2* gene has the potential to encode both transmembrane and soluble isoforms by transcriptional processing mechanisms.

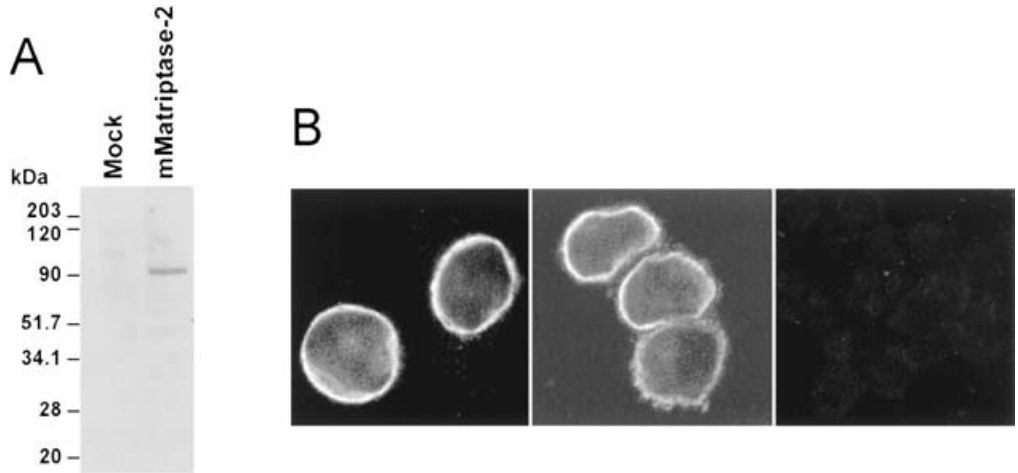


Figure 4 m-matriptase-2 expressed protein analysis

(**A**) Western-blot analysis probing with an anti-FLAG mAb of lysates from mock-transfected and m-matriptase-2-flag-transfected HeLa cells. Lysates (20 μg) were electrophoresed under reducing conditions. (**B**) Fluorescence-confocal-microscopic analysis of non-permeabilized CHO-K1 cells, either transiently transfected with a m-matriptase-2-flag expression construct (left), a m-hepsin-myc expression construct (middle) or mock-transfected (right). Cells were incubated with anti-FLAG (left and right panels) or anti-MYC (middle) mAb. Magnification 630×.

m-matriptase-2 protein characterization

Western-blot analysis was performed on lysates from mock-transfected HeLa cells and HeLa cells transiently transfected with an m-matriptase-2 expression construct containing a FLAG epitope introduced at the C-terminus of the protein. As shown in Figure 4(A), m-matriptase-2 migrated at approximately the theoretical molecular mass of 91 kDa under reducing conditions. However, under non-reducing conditions a diffuse, immunoreactive, high-molecular-mass protein band was present at the top of the separating gel (results not shown), suggesting that multimeric forms of m-matriptase-2 could be formed, possibly through the unpaired Cys^{152} discussed previously. The anti-FLAG antibody was not immunoreactive against lysates from mock-transfected HeLa cells (Figure 4A).

Immunofluorescence was used to examine the cellular location of m-matriptase-2 in transiently transfected CHO cells. Consistent with the predicted type II cell-surface orientation of m-matriptase-2, non-permeabilized CHO cells transiently transfected with the m-matriptase-2-flag construct showed strong membrane staining when incubated with an anti-FLAG mAb (Figure 4B, left panel). Hepsin has been shown to be expressed on the cell surface in a type II orientation by human hepatoma cells [50]. Accordingly, CHO cells transiently transfected with a m-hepsin expression construct encoding a MYC epitope at the C-terminus of the protein were used as a positive control for membrane localization. Similar to CHO cells expressing epitope-tagged m-matriptase-2, CHO cells expressing m-hepsin-myc showed distinct membrane staining when incubated with an anti-MYC mAb (Figure 4B, middle panel). In contrast, mock-transfected cells were essentially free of staining when incubated with either the anti-FLAG antibody (Figure 4B, right panel) or anti-MYC antibody (results not shown). These data are consistent with the predicted cell-surface location of m-matriptase-2.

m-matriptase-2 mRNA expression in adult and embryonic tissues

The distribution of m-matriptase-2 was examined in a range of adult mouse tissues by RT-PCR using m-matriptase-2-specific exon-spanning primers. Analysis of total RNA from 12 different tissues detected m-matriptase-2 at highest levels in liver (Figure 5A). m-matriptase-2 mRNA was also highly expressed in kidney and uterus. A weak, but detectable, signal for m-matriptase-2 was observed in each of the other tissues (brain, lung, heart, spleen, muscle, intestine, thymus and pancreas) analysed. As a comparison, the distribution of m-hepsin, a known liver- and kidney-expressed TTSP [32], was also assessed by RT-PCR in these tissues using m-hepsin-specific exon-spanning primers. Similar to m-matriptase-2, m-hepsin mRNA was detected at highest levels in liver and kidney. Also, similar to m-matriptase-2, m-hepsin was detected in uterus, with lower levels of expression in each of the other tissues examined. The expression of m-hepsin mRNA by mouse uterus is in contrast with the results of a previous study in which Northern-blot analysis did not detect expression in a mixed sample of ovary and uterus [32]. This difference may be due to the proportion of uterine mRNA in the mixed sample (which was not reported), to mouse strain variation, or to differences in the stage of the oestrous cycle at which the mice were killed.

In the m-hepsin RT-PCR performed on total RNA from mouse brain, in addition to the expected band at 760 bp, a second higher-molecular-mass band at approx. 850 bp was also detected (Figure 5A, arrowhead). The corresponding reaction products were cloned and sequenced. The 760-nt predicted band was confirmed to be wild-type m-hepsin RNA and the higher-molecular-mass band indicated a novel m-hepsin splice variant (deposited in the GenBank® database under accession number AY234104). A comparison of m-hepsin cDNA and genomic sequences [51] indicated that the variant transcript was generated by inclusion of the complete sequence of a 106-bp intron located between exons 7 and 8 of the reported *m-hepsin* gene [51]. The intron insertion results in an amino acid change at Cys^{137} to tryptophan and is predicted to produce a truncated version of m-hepsin completely lacking the serine-protease domain as well as 13 of the 98 residues of the m-hepsin scavenger-receptor domain (Figure 5B).

The expression of m-matriptase in adult tissues was further analysed by *in situ* hybridization, with m-hepsin serving as a comparative control on contiguous sections. Antisense and negative control sense probes used in these experiments were generated from constructs produced from the same primer pairs used in the RT-PCR analyses. Hybridizations were performed on the three tissues (liver, kidney and uterus) exhibiting the highest

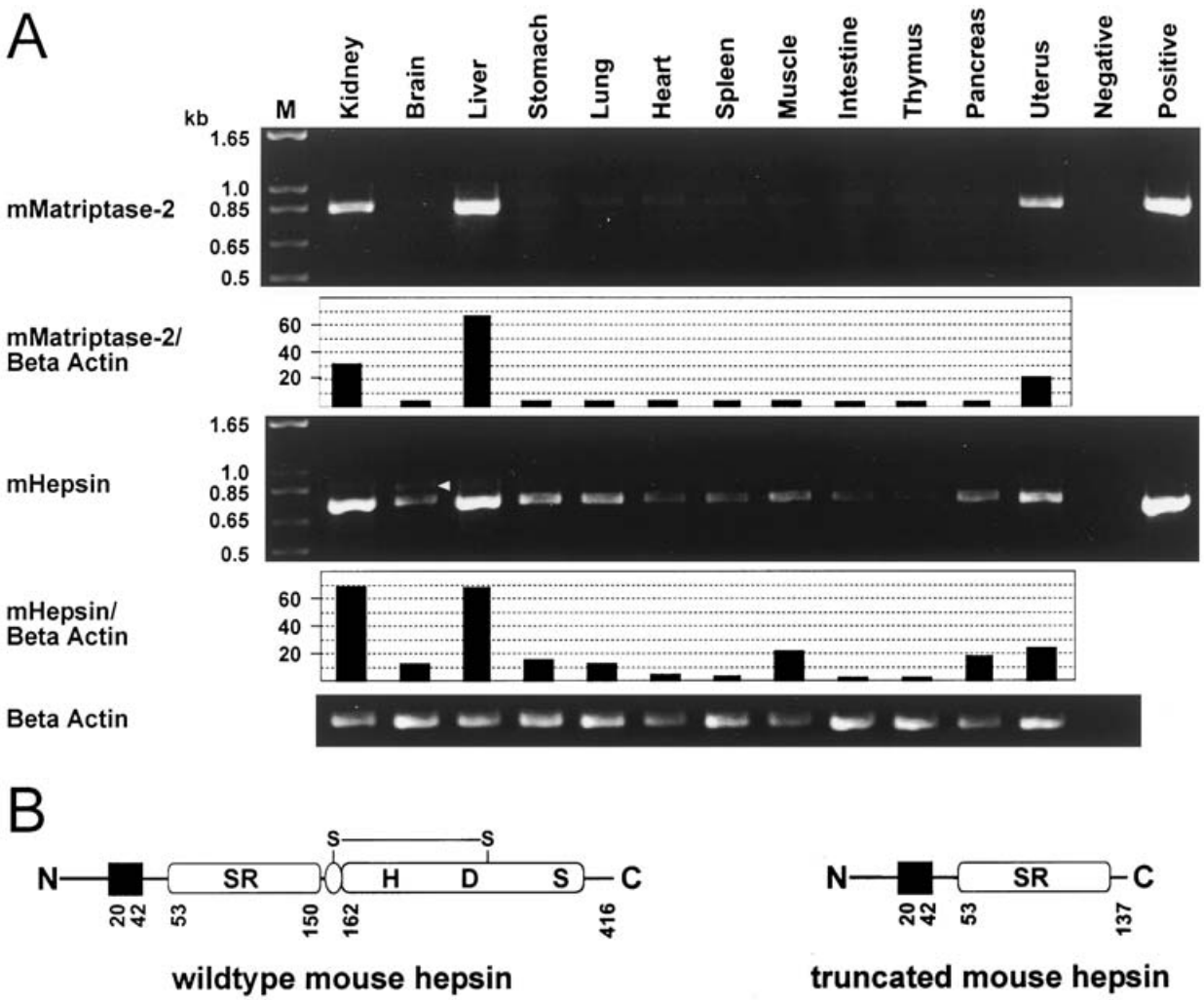


Figure 5 Distribution of m-matriptase-2 and m-hepsin in adult mouse tissues and the structure of a truncated form of m-hepsin

(**A**) RT-PCR analysis of m-matriptase-2 and m-hepsin mRNA distribution in adult mouse tissues. The 866 bp band due to a novel m-hepsin splice variant in adult brain is indicated by an arrowhead. Negative control reactions did not include template DNA. Positive control reactions included 100 pg of template DNA. The ratio of the signal intensity of m-matriptase-2 and m-hepsin relative to β-actin is displayed graphically; arbitrary units are shown. Predicted size of reaction products: m-matriptase-2, 842 bp; m-hepsin, 760 bp. (**B**) Diagrammatic representation of wild-type m-hepsin and a truncated m-hepsin protein generated by a novel splice variant expressed by adult brain. Filled box, transmembrane domain; boxed SR, scavenger-receptor domain; oval, activation domain; boxed HDS, serine-protease domain. Numbers delineate residues at the boundary of each functional domain. The disulphide bond linking pro and catalytic domains of wild-type m-hepsin is marked. The m-hepsin splice variant GenBank® accession number is AY234104.

expression of m-matriptase-2 and m-hepsin by RT-PCR analysis. Both m-matriptase-2 and m-hepsin mRNA was detectable uniformly in liver with signal associated with hepatocytes (Figure 6A and 6B). In contrast with our RT-PCR analysis, where m-hepsin was detectable at lower cycle number than m-matriptase-2, the latter probe produced a stronger signal, by *in situ* hybridization, to adult liver. m-matriptase-2 and m-hepsin mRNA were also detected in adult kidney, with the m-hepsin riboprobe showing stronger signal intensity (Figures 6C and 6G). Whereas m-matriptase-2 mRNA was detectable throughout adult kidney (Figure 6C), m-hepsin expression was predominantly associated with renal tubules of the kidney medulla (Figures 6G and 6H) with signal from the kidney cortex of much lower intensity (Figures 6F and 6H). In uterus, both m-matriptase-2 and m-hepsin were predominantly expressed by glandular columnar epithelial cells (Figures 6I and 6J). For each tissue, sense control probes showed only background levels of signal (Figures 6D, 6E, 6K and 6L; results not shown).

The distribution of m-matriptase-2 mRNA was also assessed by RT-PCR during mouse embryogenesis from 7.5 dpc to 15.5 dpc; stages at which liver and kidney organogenesis is occurring. As m-hepsin mRNA is expressed during embryogenesis [32,52] it was again used for comparative purposes. As shown in Figure 7, m-matriptase-2 mRNA was expressed at various levels during mouse embryonal development. Expression was barely detectable from day 7.5 to day 10.5, with expression at higher levels from 12.5 to 15.5 dpc with a peak in expression at day 13.5. m-hepsin mRNA was also detectable at each of these stages of development. In contrast, m-hepsin expression appeared biphasic, with peaks in expression occurring at 8.5 and 13.5 dpc.

On the basis of the increasing expression of m-matriptase-2 as assessed by RT-PCR, *in situ* hybridization was performed on mouse embryos from 12.5 to 15.5 dpc. m-hepsin was used as a comparative control on contiguous sections. Antisense and negative control sense probes used in these experiments were also generated from constructs produced from the same primer pairs

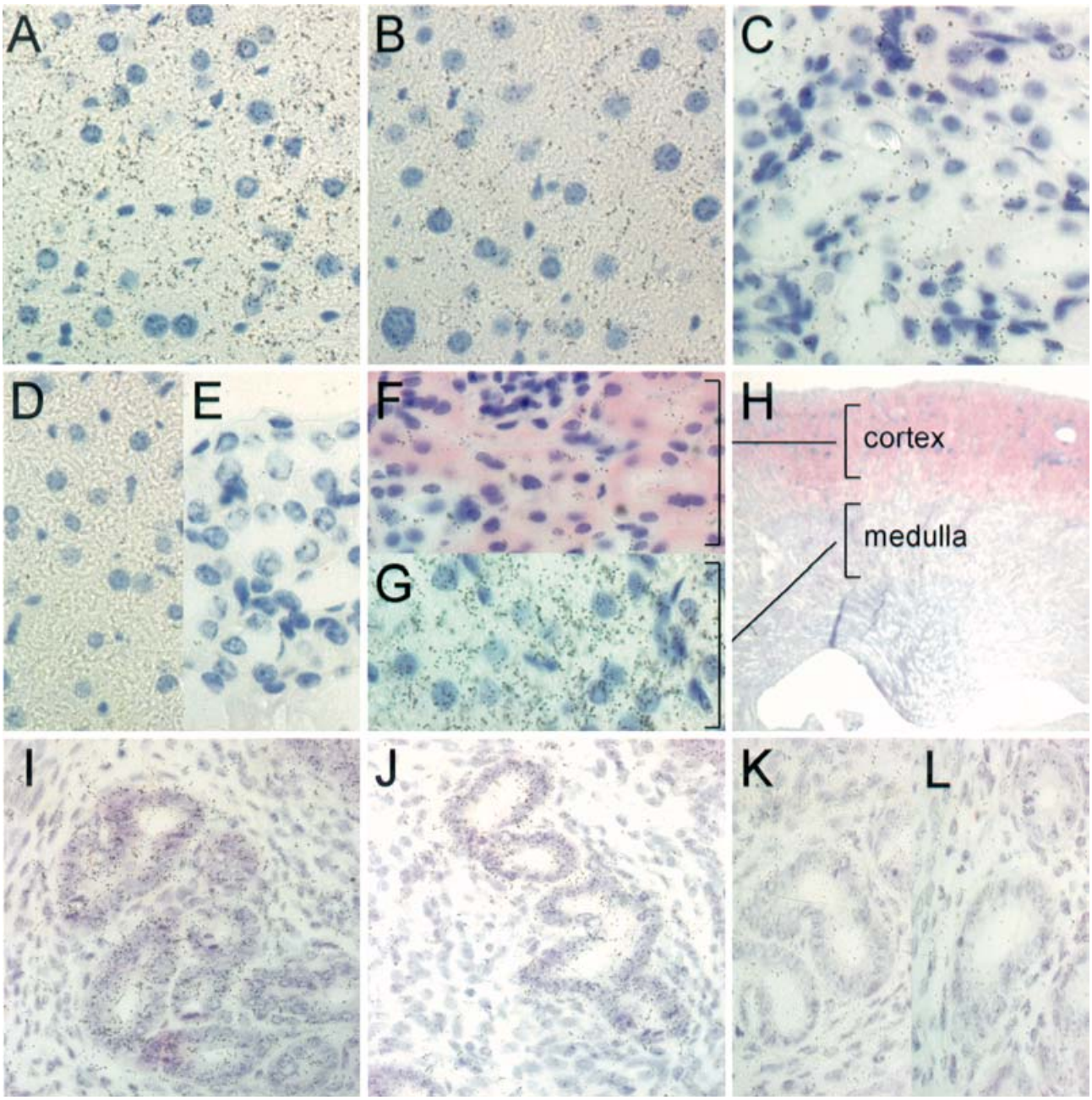


Figure 6 ***In situ*** **hybridization analysis of m-matriptase-2 and m-hepsin mRNA expression in adult tissues**

The Figure shows photomicrographs of eosin-and-haematoxylin-stained sections (5 μm thick) hybridized with antisense m-matriptase-2 (**A**), (**C** and **I**), sense m-matriptase-2 (**E** and **K**), antisense m-hepsin (**B**, **F**, **G** and **J**) or sense m-hepsin (**D**) and (**L**) ^{35}S-labelled cRNA probes. (**A**, **B** and **D**) liver; (**C**, **E**, **F** and **G**) kidney. The images in (**F**) and (**G**) were acquired from the same section of kidney. (**I**–**L**) uterus. Magnification 630×. (**H**) Eosin-and-haematoxylin-stained mouse kidney, indicating the region in which m-hepsin is predominantly expressed.

used in the RT-PCR analyses. At each stage of development the m-matriptase-2 signal was uniformly present in the developing liver (Figures 8A and 8B, arrowhead). In addition, within the nasal cavity, olfactory epithelial cells showed strong specific signal for m-matriptase-2 at each developmental stage (Figures 8A and 8B, arrow; Figure 8C), with proximal serous glands associated with the lateral wall of the nasal cavity completely clear of signal (results not shown). At 15.5 dpc pharyngo–tympanic tubes also were positive for m-matriptase-2 (Figure 8B, encircled material). In contrast with adult tissues, no signal was present in embryonic kidney. m-hepsin mRNA was also strongly expressed by cells of the developing liver from 12.5 to 15.5 dpc (Figures 8E and 8H, arrowhead). In addition, the residual embryonic yolk sac present only on day-12.5 sections showed strong signal for m-hepsin, with expression restricted to the extra-embryonic mesoderm (Figure 8E, asterisk; Figure 8F). In contrast the yolk sac present in Figure 8A (asterisk) showed only background levels of signal for m-matriptase-2. Consistent with a previous report of immunohistochemical localization of m-hepsin in embryonic tissues [52], at 15.5 dpc, m-hepsin expression appeared more widespread, with close analysis of the section indicating specific signal from lung, heart, gastrointestinal tract, developing kidney

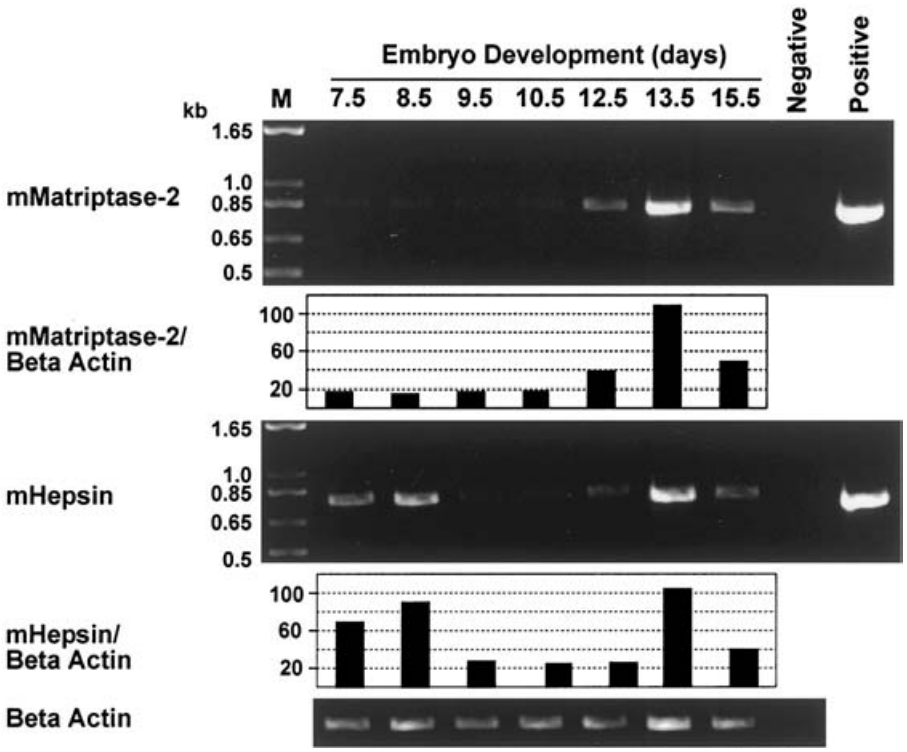


Figure 7 Distribution of m-matriptase-2 and m-hepsin during mouse embryogenesis

RT-PCR analysis of m-matriptase-2 and m-hepsin mRNA in day 7.5–15.5 embryos. Negative control reactions did not include template DNA. Positive control reactions included 100 pg of template DNA. The ratio of the signal intensity of m-matriptase-2 and m-hepsin relative to β-actin is displayed graphically; arbitrary units are shown. Predicted size of reaction products: m-matriptase-2, 842 bp; m-hepsin, 760 bp.

tubules and epithelium of the oral cavity (Figure 8H). As was observed for m-matriptase-2, pharyngo–tympanic tubes also were positive for m-hepsin expression at 15.5 dpc (Figure 8H, encircled material), Sense control probes showed only background levels of signal (Figures 8D and 8G; results not shown).

DISCUSSION

Here we report the identification of the type II transmembrane serine protease m-matriptase-2, a 91 kDa, 811-amino-acid protein consisting of a 59-residue cytoplasmic tail, two CUB domains, three LDLR domains and a C-terminal trypsin-like serine-protease domain. The putative membrane location of m-matriptase-2 is supported by immunofluorescence localization of the protein and is consistent with the previously described type II orientation at the plasma membrane of h-matriptase-2 [39]. The m-hepsin immunofluorescence analysis described here was also consistent with the reported cell-surface localization of human hepsin [50]. m-matriptase-2 is predicted to be a pro-enzyme with cleavage required between Arg^{576} and Ile^{577} for enzyme activation. As for each of the other members of the TTSP family [1], following activation the proteolytic domain is expected to remain tethered to the activation domain by a highly conserved disulphide bond.

m-matriptase-2, r-matriptase-2 and h-matriptase-2 share a high degree of sequence and structural identity. Each contains two CUB, three LDLR, protease-activation and serine-protease domains, as well as conserved consensus N-glycosylation sites and cysteine residues. The cysteine residues are predicted to form 18 intramolecular disulphide bonds, with one unpaired cysteine residue preceding the first CUB domain available for potential homo- and hetero-intermolecular linkages. Interestingly, whereas the m-matriptase-2 and h-matriptase-2 span the plasma membrane, our genomic sequence analysis indicated that the *r-matriptase*-2 gene, through use of alternate exons, will be capable of producing both a membrane-spanning protein as well as a secreted isoform. Accordingly the generation of transmembrane and soluble isoforms of r-matriptase-2 is predicted to occur via transcriptional processing mechanisms, as has been reported for the TTSP family member MSPL [23]. In contrast, analysis of *h-matriptase-2* and *m-matriptase-2* genomic sequence did not identify alternate *r-matriptase-2*-like signal-peptide-encoding exons (results not shown). Accordingly, it appears likely that, if they exist, generation of soluble forms of h- and m-matriptase-2 will require post-translational processing events such as have been demonstrated for another member of the TTSP family m-matriptase [36] rather than transcriptional mechanisms. In addition to variation at the N-terminus, r-matriptase-2 contains an additional 17 amino acids following the first CUB domain. It is not yet clear whether this insertion will have any species-specific effect on r-matriptase-2 function relative to the human and mouse orthologues. Indeed it will be of interest to determine whether human, mouse and rat membrane bound matriptase-2 isoforms have orthologous functions and whether soluble forms of this protein are produced in each species and also to elucidate the mechanisms at the gene and protein level which regulate generation of soluble isoforms.

The genes encoding h- and m-matriptase-2 are located at syntenic chromosome positions (22q13.1 and 15E2 respectively). In addition, these genes share identical structures spanning 18 exons with the human gene encompassing approx. 38 kb of

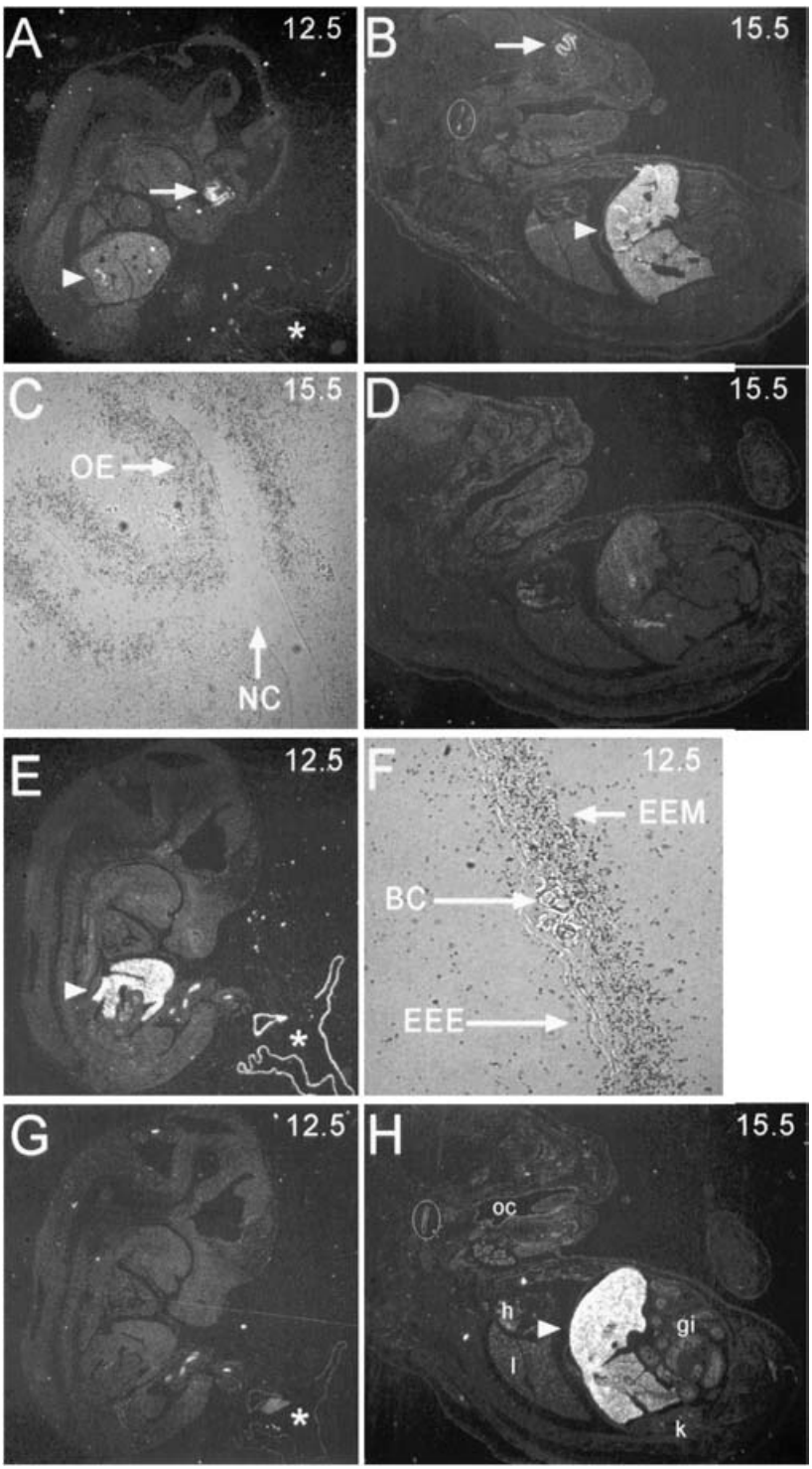


Figure 8 ***In-situ*-hybridization analysis of m-matriptase-2 and m-hepsin mRNA expression in day-12.5 and day-15.5 embryos**

The Figure shows dark- and bright-field photomicrographs of sections (5 μm thick) hybridized with antisense m-matriptase-2 (**A–C**), sense m-matriptase-2 (**D**), antisense m-hepsin (**E**, **F** and **H**) or sense m-hepsin (**G**) ^{35}S-labelled cRNA probes. (**A**, **E** and **G**) 12.5 dpc mouse embryo; (**B**, **D** and **H**) 15.5 dpc embryo. Arrowhead, embryonic liver; arrow, nasal cavity; circle, pharyngo–tympanic tubes; asterisk, day-12.5 yolk sac; oc, oral cavity; l, lung; h, heart; gi, gastrointestinal tract; k, kidney. (**C**) Higher magnification of nasal cavity of day 15.5 embryo. OE, olfactory epithelium; NC, nasal cavity. (**F**) Higher magnification of yolk sac of day 12.5 embryo. BC, red blood cells; EEE, extra embryonic ectoderm; EEM, extra embryonic mesoderm. (**A**, **B**, **D**, **E**, **G** and **H**): magnification 10.6×; (**C** and **F**): magnification 630×.

genomic sequence (see our GenBank® entry AY055383) and m-matriptase-2 approx. 29 kb of sequence. The correspondence of m-matriptase-2 protein structural domains with intron/exon boundaries of the encoding gene may be a general feature of TTSP protein/genomic structure. This correspondence is shared by human [25] and mouse [33] TMPRSS2, m-hepsin [51], human and

mouse corin [53] and human enteropeptidase [54], and supports the proposal by Pan et al. [53] that the genes encoding human and mouse corin and other TTSP family members arose from duplication and rearrangement of pre-existing exons encoding structurally distinct domains.

In adult tissues, m-matriptase-2 mRNA is expressed at highest levels in liver, kidney and uterus and at much lower levels in each of the other nine tissues analysed by RT-PCR. In contrast, it has recently been reported that, by Northern-blot analysis of 16 adult human tissues, h-matriptase-2 mRNA was detected exclusively in liver [39]. This species difference in tissue distribution likely indicates differences in transcriptional regulation of *matriptase-2* gene expression in human and mouse. It also parallels the differences in expression patterns of hepsin in mouse and primate species. Our analysis indicated that m-hepsin mRNA is also expressed at highest levels in liver, kidney and uterus, which contrasts with the initial reports characterizing hepsin which reported high levels of mRNA expression only in liver in primate species [50]. The similarity in expression patterns of matriptase-2 and hepsin observed in mouse adult tissues by RT-PCR analysis was also apparent in mouse embryonic development. Between 7.5 and 15.5 dpc the expression of both TTSPs was apparent at each time point analysed, with a peak in expression at 13.5 dpc. In addition, m-hepsin expression showed a second expression peak at 8.5 dpc. The overlapping expression profiles of matriptase-2 and hepsin in human and mouse in adult tissues and during embryogenesis likely indicates that the encoding genes share at least some transcriptional regulatory elements.

Using *in situ* hybridization we were able to determine that both m-matriptase-2 and m-hepsin are expressed uniformly throughout adult liver by hepatocytes. However, in adult kidney, while m-matriptase-2 was distributed uniformly throughout this organ, m-hepsin expression was more restricted and predominantly associated with renal tubules of the kidney medulla. Interestingly, immunohistochemical analysis of normal human tissues using rabbit polyclonal antibodies failed to detect hepsin in human adult kidney [55]. It is not yet known whether this difference is due to the species-specific expression pattern of hepsin or whether in mouse kidney the gene is transcribed but not translated. In uterus, both TTSPs were expressed predominantly by glandular epithelial cells. In embryonic tissues both m-matriptase-2 and m-hepsin were strongly expressed in the developing liver. The genes encoding both these proteins were also expressed during embryogenesis by epithelial cells of developing pharyngo–tympanic tubes. However, as in adult tissues, although RT-PCR analysis indicated very similar gross tissue expression patterns for m-matriptase-2 and m-hepsin, distinct differences were also apparent in embryonic tissues by *in situ* hybridization. For example, while m-matriptase-2 was strongly expressed within the nasal cavity by olfactory epithelial cells, m-hepsin was detected in the residual embryonic yolk sac and with lower intensity in lung, heart, gastrointestinal tract, developing kidney tubules and epithelium of the oral cavity. Accordingly, although m-matriptase-2 and m-hepsin share similarities in expression patterns, likely indicating shared transcriptional regulatory mechanisms, it is apparent that the expression patterns of the encoding genes do not completely overlap.

The identification of a novel brain m-hepsin splice variant is further evidence of differences in the regulation of m-matriptase-2 and m-hepsin expression. While four other m-hepsin splice variants have been reported [32,51], our RT-PCR analysis detected only one m-matriptase-2 mRNA. Of course as the m-matriptase-2 primers used in our analysis spanned only 842 bp of the $\approx$3.2 kb cDNA, it is possible that other m-matriptase-2 splice variants will exist. It will be of interest to determine whether other matriptase-2 splice variants occur in human and mouse, and whether these, or the reported hepsin splice variants, will encode functional proteins. In this respect, the protein deduced from the novel m-hepsin mRNA reported here contains a short cytoplasmic tail, a membrane-spanning region and most of a scavenger-receptor domain, therefore retaining the structural characteristics necessary to function at the cell surface as a potential transmembrane receptor for soluble, cell-surface or matrix ligands.

In summary, among the TTSPs, m-matriptase-2 exhibits a novel expression profile that overlaps with the distribution of m-hepsin. A functional role in several cell types is indicated by *in situ* hybridization analysis, which localized m-matriptase-2 expression in adult tissues to liver hepatocytes and to uterine epithelial cells, with uniform expression in kidney. In addition, a role in mouse development was suggested by the expression of m-matriptase in the developing liver as well as by a restricted set of embryonic epithelial cells, including epithelial cells of the nasal cavity and pharyngo–tympanic tubes. Two members of the TTSP family have demonstrated roles as protein/peptide activators. One of these, the heart-expressed enzyme corin, likely functions in an autocrine-type activation process whereby pro-ANP produced by cardiomyocytes is converted on the surface of these cells into biologically active ANP by corin [5,6]. The other TTSP, the duodenum protease enteropeptidase, functions in an endocrine-type activation process whereby trypsinogen, secreted by the pancreas into the duodenum, is catalytically activated to trypsin by the direct action of enteropeptidase [2]. As m-matriptase-2 has the structural features necessary to function as a serine protease, it is likely that it, too, will have a role as a protein/peptide activator or convertase. Whether this is via an autocrine-type or endocrine-type process is not known. However, the expression of m-matriptase-2 by epithelial cells undergoing tubular morphogenesis in the embryo and requiring rapid renewal in the adult, as well as by highly metabolic cells of the liver and kidney, is suggestive that this enzyme is more likely to function in an autocrine-type activation process. It should be noted that a functional role or substrate has been defined for very few of the members of the TTSP family [56]. On the basis of the recent report of matriptase/MT-SP1$^{-/-}$ mice which demonstrated that this TTSP functions in the development of the epidermis and thymic homoeostasis [21], it is apparent that a major experimental route to determine the physiological function and natural substrates of these new serine proteases is through gene-knockout approaches. Accordingly deletion of the *matriptase-2* gene in mouse should assist in defining the function of this newest member of the TTSP family.

Although most of the TTSPs exhibit abundant mRNA expression in a restricted number of tissues, there is considerable overlap in tissue distribution patterns among the TTSP family members when the lower-expressing tissues are taken into account [1]. It is not known whether overlapping expression also indicates some extent of functional redundancy amongst family members. However, one of the two groups that have reported hepsin$^{-/-}$ mice has proposed that the lack of gross phenotypic differences between null mice and wild-type controls may be due to compensation by a functionally related gene product [57]. Indeed, on the basis of previous reports of potential hepsin involvement in blastocyst hatching [32], cell growth [52] and blood coagulation [58], and in contrast with the effect in mouse of deletion of matriptase/MT-SP1 [21], it was surprising that the only gross phenotype of hepsin$^{-/-}$ mice was elevated levels of bone-derived alkaline phosphatase [57,59]. Accordingly, on the basis of the similar expression patterns of m-matriptase-2 and m-hepsin it is possible, at least in mouse, that these enzymes share a level

of functional redundancy sufficient to overcome the phenotypic changes expected of, but not observed, in hepsin$^{-/-}$ mice. The proposal of some level of functional redundancy between m-matriptase-2 and m-hepsin could be tested by generation of double-knockout hepsin/matriptase-2 mice.

We thank Dr Michael Fitch, The Scripps Research Institute, for embryonic mouse RNA and for helpful discussions, and Dr Ronald Aimes and Dr Andries Zijlstra for helpful discussions. This work was supported by National Institutes of Health Grants CA55852 and HL31950 (J. P. Q.), HL65738 (H. S.) and training grant T32 HL07695 (G. G.) and a National Health and Medical Research Council of Australia CJ Martin/RG Menzies Fellowship 138722 (J. D. H.). This is The Scripps Research Institute Manuscript Number 15615-CB.

REFERENCES

1 Hooper, J. D., Clements, J. A., Quigley, J. P. and Antalis, T. M. (2001) Type II transmembrane serine proteases. Insights into an emerging class of cell surface proteolytic enzymes. J. Biol. Chem. **276**, 857–860
2 Kitamoto, Y., Veile, R. A., Donis-Keller, H. and Sadler, J. E. (1995) cDNA sequence and chromosomal localization of human enterokinase, the proteolytic activator of trypsinogen. Biochemistry **34**, 4562–4568
3 Yan, W., Sheng, N., Seto, M., Morser, J. and Wu, Q. (1999) Corin, a mosaic transmembrane serine protease encoded by a novel cDNA from human heart. J. Biol. Chem. **274**, 14926–14935
4 Hooper, J. D., Scarman, A. L., Clarke, B. E., Normyle, J. F. and Antalis, T. M. (2000) Localization of the mosaic transmembrane serine protease corin to heart myocytes. Eur. J. Biochem. **267**, 6931–6937
5 Yan, W., Wu, F., Morser, J. and Wu, Q. (2000) Corin, a transmembrane cardiac serine protease, acts as a pro-atrial natriuretic peptide-converting enzyme. Proc. Natl. Acad. Sci. U.S.A. **97**, 8525–8529
6 Wu, F., Yan, W., Pan, J., Morser, J. and Wu, Q. (2002) Processing of pro-atrial natriuretic peptide by corin in cardiac myocytes. J. Biol. Chem. **277**, 16900–16905
7 Luo, J., Duggan, D. J., Chen, Y., Sauvageot, J., Ewing, C. M., Bittner, M. L., Trent, J. M. and Isaacs, W. B. (2001) Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. Cancer Res. **61**, 4683–4688
8 Magee, J. A., Araki, T., Patil, S., Ehrig, T., True, L., Humphrey, P. A., Catalona, W. J., Watson, M. A. and Milbrandt, J. (2001) Expression profiling reveals hepsin overexpression in prostate cancer. Cancer Res. **61**, 5692–5696
9 Welsh, J. B., Sapinoso, L. M., Su, A. I., Kern, S. G., Wang-Rodriguez, J., Moskaluk, C. A., Frierson, Jr, H. F. and Hampton, G. M. (2001) Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. Cancer Res. **61**, 5974–5978
10 Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A. and Chinnaiyan, A. M. (2001) Delineation of prognostic biomarkers in prostate cancer. Nature (London) **412**, 822–826
11 Stamey, T. A., Warrington, J. A., Caldwell, M. C., Chen, Z., Fan, Z., Mahadevappa, M., McNeal, J. E., Nolley, R. and Zhang, Z. (2001) Molecular genetic profiling of Gleason grade 4/5 prostate cancers compared to benign prostatic hyperplasia. J. Urol. **166**, 2171–2177
12 Ernst, T., Hergenhahn, M., Kenzelmann, M., Cohen, C. D., Bonrouhi, M., Weninger, A., Klaren, R., Grone, E. F., Wiesel, M., Gudemann, C. et al. (2002) Decrease and gain of gene expression are equally discriminatory markers for prostate carcinoma – A gene expression analysis on total and microdissected prostate tissue. Am. J. Pathol. **160**, 2169–2180
13 Lin, C.-Y., Anders, J., Johnson, M., Sang, Q. A. and Dickson, R. B. (1999) Molecular cloning of cDNA for matriptase, a matrix-degrading serine protease with trypsin-like activity. J. Biol. Chem. **274**, 18231–18236
14 Takeuchi, T., Shuman, M. A. and Craik, C. S. (1999) Reverse biochemistry: use of macromolecular protease inhibitors to dissect complex biological processes and identify a membrane-type serine protease in epithelial cancer and normal tissue. Proc. Natl. Acad. Sci. U.S.A. **96**, 11054–11061
15 Oberst, M., Anders, J., Xie, B., Singh, B., Ossandon, M., Johnson, M., Dickson, R. B. and Lin, C. Y. (2001) Matriptase and HAI-1 are expressed by normal and malignant epithelial cells *in vitro* and *in vivo*. Am. J. Pathol. **158**, 1301–1311
16 Oberst, M. D., Johnson, M. D., Dickson, R. B., Lin, C. Y., Singh, B., Stewart, M., Williams, A., al Nafussi, A., Smyth, J. F., Gabra, H. and Sellar, G. C. (2002) Expression of the serine protease matriptase and its inhibitor HAI-1 in epithelial ovarian cancer: correlation with clinical outcome and tumor clinicopathological parameters. Clin. Cancer Res. **8**, 1101–1107
17 Tanimoto, H., Underwood, L. J., Wang, Y. X., Shigemasa, K., Parmley, T. H. and O'Brien, T. J. (2001) Ovarian tumor cells express a transmembrane serine protease: a potential candidate for early diagnosis and therapeutic intervention. Tumor Biol. **22**, 104–114
18 Takeuchi, T., Harris, J., Huang, W., Yan, K. W., Coughlin, S. R. and Craik, C. S. (2000) Cellular localization of membrane-type serine protease 1 and identification of protease activated receptor-2 and single-chain urokinase-type plasminogen activator as substrates. J. Biol. Chem. **275**, 26333–26342
19 Lee, S. L., Dickson, R. B. and Lin, C. Y. (2000) Activation of hepatocyte growth factor and urokinase/plasminogen activator by matriptase, an epithelial membrane serine protease. J. Biol. Chem. **275**, 36720–36725
20 Friedrich, R., Fuentes-Prior, P., Ong, E., Coombs, G., Hunter, M., Oehler, R., Pierson, D., Gonzalez, R., Huber, R., Bode, W. and Madison, E. L. (2002) Catalytic domain structures of MT-SP1/matriptase, a matrix-degrading transmembrane serine proteinase. J. Biol. Chem. **277**, 2160–2168
21 List, K., Haudenschild, C. C., Szabo, R., Chen, W., Wahl, S. M., Swaim, W., Engelholm, L. H., Behrendt, N. and Bugge, T. H. (2002) Matriptase/MT-SP1 is required for postnatal survival, epidermal barrier function, hair follicle development and thymic homeostasis. Oncogene **21**, 3765–3779
22 Yamaoka, K., Masuda, K., Ogawa, H., Takagi, K., Umemoto, N. and Yasuoka, S. (1998) Cloning and characterization of the cDNA for human airway trypsin-like protease. J. Biol. Chem. **273**, 11895–11901
23 Kim, D. R., Sharmin, S., Inoue, M. and Kido, H. (2001) Cloning and expression of novel mosaic serine proteases with and without a transmembrane domain from human lung. Biochim. Biophys. Acta **1518**, 204–209
24 Lang, J. C. and Schuller, D. E. (2001) Differential expression of a novel serine protease homologue in squamous cell carcinoma of the head and neck. Br. J. Cancer **84**, 237–243
25 Paoloni-Giacobino, A., Chen, H., Peitsch, M. C., Rossier, C. and Antonarakis, S. E. (1997) Cloning of the TMPRSS2 gene, which encodes a novel serine protease with transmembrane, LDLRA and SRCR domains and maps to 21q22.3. Genomics **44**, 309–320
26 Scott, H. S., Kudoh, J., Wattenhofer, M., Shibuya, K., Berry, A., Chrast, R., Guipponi, M., Wang, J., Kawasaki, K., Asakawa, S. et al. (2001) Insertion of β-satellite repeats identifies a transmembrane protease causing both congenital and childhood onset autosomal recessive deafness. Nat. Genet. **27**, 59–63
27 Wallrapp, C., Hahnel, S., Muller-Pillasch, F., Burghardt, B., Iwamura, T., Ruthenburger, M., Lerch, M. M., Adler, G. and Gress, T. M. (2000) A novel transmembrane serine protease (TMPRSS3) overexpressed in pancreatic cancer. Cancer Res. **60**, 2602–2606
28 Yamaguchi, N., Okui, A., Yamada, T., Nakazato, H. and Mitsui, S. (2001) Spinesin/TMPRSS5, a novel transmembrane serine protease, cloned from human spinal cord. J. Biol. Chem. **277**, 6806–6812
29 Yuan, X., Zheng, X., Lu, D., Rubin, D. C., Pung, C. Y. and Sadler, J. E. (1998) Structure of murine enterokinase (enteropeptidase) and expression in small intestine during development. Am. J. Physiol. **274**, 342–349
30 Kim, M. G., Chen, C., Lyu, M. S., Cho, E. G., Park, D., Kozak, C. and Schwartz, R. H. (1999) Cloning and chromosomal mapping of a gene isolated from thymic stromal cells encoding a new mouse type II membrane serine protease, epithin, containing four LDL receptor modules and two CUB domains. Immunogenetics **49**, 420–428
31 Tomita, Y., Kim, D. H., Magoori, K., Fujino, T. and Yamamoto, T. T. (1998) A novel low-density lipoprotein receptor-related protein with type II membrane protein-like structure is abundant in heart. J. Biochem. (Tokyo) **124**, 784–789
32 Vu, T. K. H., Liu, R. W., Haaksma, C. J., Tomasek, J. J. and Howard, E. W. (1997) Identification and cloning of the membrane-associated serine protease, hepsin, from mouse preimplantation embryos. J. Biol. Chem. **272**, 31315–31320
33 Jacquinet, E., Rao, N. V., Rao, G. V. and Hoidal, J. R. (2000) Cloning, genomic organization, chromosomal assignment and expression of a novel mosaic serine proteinase: epitheliasin. FEBS Lett. **468**, 93–100
34 Lin, C.-Y., Wang, J. K., Torri, J., Dou, L., Sang, Q. A. and Dickson, R. B. (1997) Characterization of a novel, membrane-bound, 80-kDa matrix-degrading protease from human breast cancer cells. Monoclonal antibody production, isolation and localization. J. Biol. Chem. **272**, 9147–9152
35 Lin, C.-Y., Anders, J., Johnson, M. and Dickson, R. B. (1999) Purification and characterization of a complex containing matriptase and a Kunitz-type serine protease inhibitor from human milk. J. Biol. Chem. **274**, 18237–18242
36 Cho, E. G., Kim, M. G., Kim, C., Kim, S. R., Seong, I. S., Chung, C., Schwartz, R. H. and Park, D. (2001) N-terminal processing is essential for release of epithin, a mouse type II membrane serine protease. J. Biol. Chem. **276**, 44581–44589
37 Fonseca, P. and Light, A. (1983) The purification and characterization of bovine enterokinase from membrane fragments in the duodenal mucosal fluid. J. Biol. Chem. **258**, 14516–14520

38 Afar, D. E., Vivanco, I., Hubert, R. S., Kuo, J., Chen, E., Saffran, D. C., Raitano, A. B. and Jakobovits, A. (2001) Catalytic cleavage of the androgen-regulated TMPRSS2 protease results in its secretion by prostate and prostate cancer epithelia. Cancer Res. **61**, 1686–1692
39 Velasco, G., Cal, S., Quesada, V., Sanchez, L. M. and Lopez-Otin, C. (2002) Matriptase-2, a membrane-bound mosaic serine proteinase predominantly expressed in human liver and showing degrading activity against extracellular matrix proteins. J. Biol. Chem. **277**, 37637–37646
40 Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C. J., Hofmann, K. and Bairoch, A. (2002) The PROSITE database, its status in 2002. Nucleic Acids Res. **30**, 235–238
41 Schultz, J., Milpetz, F., Bork, P. and Ponting, C. P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. U.S.A. **95**, 5857–5864
42 Nakai, K. and Kanehisa, M. (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics **14**, 897–911
43 Blom, N., Gammeltoft, S. and Brunak, S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol. **294**, 1351–1362
44 Sassoon, D. and Rosenthal, N. (1993) Detection of messenger RNA by *in situ* hybridization. Methods Enzymol. **225**, 384–404
45 Xiong, J. W., Leahy, A., Lee, H. H. and Stuhlmann, H. (1999) Vezf1: A Zn finger transcription factor restricted to endothelial cells and their precursors. Dev. Biol. **206**, 123–141
46 Shapiro, M. B. and Senapathy, P. (1987) RNA splice junctions of different classes of eukaryotes – sequence statistics and functional implications in gene expression. Nucleic Acids Res. **15**, 7155–7174
47 Varela, P. F., Romero, A., Sanz, L., Romao, M. J., Topfer-Petersen, E. and Calvete, J. J. (1997) The 2.4 angstrom resolution crystal structure of boar seminal plasma PSP-I/PSP-II: a zona pellucida-binding glycoprotein heterodimer of the spermadhesin family built by a CUB domain architecture. J. Mol. Biol. **274**, 635–649
48 Rudenko, G., Henry, L., Henderson, K., Ichtchenko, K., Brown, M. S., Goldstein, J. L. and Deisenhofer, J. (2002) Structure of the LDL receptor extracellular domain at endosomal pH. Science **298**, 2353–2358
49 von Heijne, G. (1985) Signal sequences. The limits of variation. J. Mol. Biol. **184**, 99–105
50 Tsuji, A., Torres-Rosado, A., Arai, T., Le Beau, M. M., Lemons, R. S., Chou, S. H. and Kurachi, K. (1991) Hepsin, a cell membrane-associated protease. Characterization, tissue distribution and gene localization. J. Biol. Chem. **266**, 16948–16953
51 Kawamura, S., Kurachi, S., Deyashiki, Y. and Kurachi, K. (1999) Complete nucleotide sequence, origin of isoform and functional characterization of the mouse hepsin gene. Eur. J. Biochem. **262**, 755–764
52 Torres-Rosado, A., O'Shea, K. S., Tsuji, A., Chou, S. H. and Kurachi, K. (1993) Hepsin, a putative cell-surface serine protease, is required for mammalian cell growth. Proc. Natl. Acad. Sci. U.S.A. **90**, 7181–7185
53 Pan, J., Hinzmann, B., Yan, W., Wu, F., Morser, J. and Wu, Q. (2002) Human and murine corin genes: Genomic structures and functional GATA elements in their promoters. J. Biol. Chem. **277**, 38390–38398
54 Holzinger, A., Maier, E. M., Buck, C., Mayerhofer, P. U., Kappler, M., Haworth, J. C., Moroz, S. P., Hadorn, H. B., Sadler, J. E. and Roscher, A. A. (2002) Mutations in the proenteropeptidase gene are the molecular cause of congenital enteropeptidase deficiency. Am. J. Hum. Genet. **70**, 20–25
55 Zacharski, L. R., Ornstein, D. L., Memoli, V. A., Rousseau, S. M. and Kisiel, W. (1998) Expression of the factor VII activating protease, hepsin, *in situ* in renal cell carcinoma. Thromb. Haemostasis **79**, 876–877
56 Netzel-Arnett, S., Hooper, J. D., Szabo, R., Madison, E. L., Quigley, J. P., Bugge, T. H. and Antalis, T. M. (2003) Membrane anchored serine proteases: an emerging class of cell surface proteolytic enzymes with potential roles in cancer. Cancer Metastasis Rev. **22**, 237–258
57 Yu, I. S., Chen, H. J., Lee, Y. S. E., Huang, P. H., Lin, S. R., Tsai, T. W. and Lin, S. W. (2000) Mice deficient in hepsin, a serine protease, exhibit normal embryogenesis and unchanged hepatocyte regeneration ability. Thromb. Haemostasis **84**, 865–870
58 Kazama, Y., Hamamoto, T., Foster, D. C. and Kisiel, W. (1995) Hepsin, a putative membrane-associated serine protease, activates human factor VII and initiates a pathway of blood coagulation on the cell surface leading to thrombin formation. J. Biol. Chem. **270**, 66–72
59 Wu, Q., Yu, D., Post, J., Halks Miller, M., Sadler, J. E. and Morser, J. (1998) Generation and characterization of mice deficient in hepsin, a hepatic transmembrane serine protease. J. Clin. Invest. **101**, 321–326

Received 11 March 2003/24 April 2003; accepted 13 May 2003
Published as BJ Immediate Publication 13 May 2003, DOI 10.1042/BJ20030390